

PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of

Docket No: CA1122

Yihong GONG, et al.

Appln. No.: 09/817,591

Group Art Unit: 2176

Confirmation No.: 7751

Examiner: Quoc A. TRAN

Filed: March 26, 2001

For: TEXT SUMMARIZATION USING RELEVANCE MEASURES AND LATENT
SEMANTIC ANALYSIS

SUBMISSION OF APPEAL BRIEF

MAIL STOP APPEAL BRIEF - PATENTS

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith please find an Appeal Brief. The USPTO is directed and authorized to charge the statutory fee of \$540.00 and all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account.

Respectfully submitted,



Fariba Sirjani
Registration No. 47,947

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON DC OFFICE

23373

CUSTOMER NUMBER

Date: April 3, 2009

PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re application of

Docket No: CA1122

Yihong GONG, et al.

Appln. No.: 09/817,591

Group Art Unit: 2176

Confirmation No.: 7751

Examiner: Quoc A. TRAN

Filed: March 26, 2001

For: TEXT SUMMARIZATION USING RELEVANCE MEASURES AND LATENT
SEMANTIC ANALYSIS

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

MAIL STOP APPEAL BRIEF - PATENTS

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 41.37, Appellant submits the following:

Table of Contents

I.	REAL PARTY IN INTEREST	2
II.	RELATED APPEALS AND INTERFERENCES.....	3
III.	STATUS OF CLAIMS	4
IV.	STATUS OF AMENDMENTS	5
V.	SUMMARY OF THE CLAIMED SUBJECT MATTER	6
VI.	GROUND OF REJECTION TO BE REVIEWED ON APPEAL	12
VII.	ARGUMENT	13
	CLAIMS APPENDIX.....	28
	EVIDENCE APPENDIX:.....	34
	RELATED PROCEEDINGS APPENDIX	35

I. REAL PARTY IN INTEREST

Based on the information supplied by the Appellant, and to the best of knowledge of the legal representative of the Appellant, the real party in interest in this Appeal is the Assignee, NEC USA, INC., C&C Research Laboratories, 4 Independence Way, Princeton, New Jersey 08540. The Assignment was previously submitted and recorded on July 16, 2001, at Reel 011989, Frame 0465.

II. RELATED APPEALS AND INTERFERENCES

None.

To the knowledge and belief of the Appellant, the Assignee, and the Appellant's legal representative, there are no other appeals or interferences before the Board of Appeals and Interferences that will directly affect or be affected by the Board's decision in the instant Appeal.

A Notice of Appeal for the present Application was filed on February 3, 2009.

III. STATUS OF CLAIMS

Claims 1-32 are pending in the Application.

Claims 1-32 are rejected.

Claims 1-20 are on Appeal.

The claims under Appeal are rejected under 35 U.S.C. 103(a) as allegedly unpatentable: claims 1-4 and 9-16 over Hwang (U.S. Patent Application Publication No. 2002/0078090 A1) in view of Foltz (U.S. Patent No. 6,356,864); claims 5, 7, 17 and 19 over Hwang, in view of Foltz, further in view of Boguraev (U.S. Patent No. 6,865,572); and claims 6, 8, 18, and 20 over Hwang, in view of Foltz, further in view of Boguraev, and further in view of Cohen (William W. Cohen, "Data Integration Using Similarity Joins and a Word-Based Information Representation Language," by AT&T Lab-Research Shannon Laboratory, Published by ACM, Vol. 18; No. 3, July 200, pages 288-321.)

A copy of the claims on Appeal is set forth in the attached Claims Appendix.

IV. STATUS OF AMENDMENTS

All amendments to the claims have been entered and are of record.

The Appellant is not seeking and has not sought to amend the claims after the amendments and response that was filed on June 4, 2008, to the non-final rejection of March 4, 2008. Consequently, the response filed on December 4, 2008, to the final Office action of September 4, 2008, did not include any claim amendments.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

Claims 1-20 are under Appeal. Claims 1, 9 and 13 are independent. The paragraph numbers used below correspond to the published version of the Application. (Gong, U.S. Patent Application Publication No. 2002/0138528.)

The present Application is directed to methods and systems for providing a generic text summary. (Specification, p. 4, lines 19-21; paragraph [0015].) As an alternative to the generic method of text summarization, and to provide context for the invention, the specification also discusses query-relevant text summarization. (Specification, p. 4, lines 13-19; paragraph [0015].) Query-relevant summaries, that are NOT the subject of the present invention, are generated in response to a user query. (Specification, p. 2, lines 15-18; paragraph [0009].) Query-relevant summaries essentially retrieve query-relevant sentences from the document to generate a summary that is specifically related to the initial search query. Consequently, query-relevant summaries may not provide an overall sense of the document content. (Specification, p. 4, lines 13-19; paragraph [0015].) On the other hand, generic text summaries provide an overall sense of the document content in general. A good generic summary should contain all the main topics presented in a document while minimizing redundancy regarding the dominant topic. (Specification, p. 2, lines 11-12; paragraph [0009].)

The following paragraph of the specification explains the issues associated with forming a generic text summary:

[0023] By way of background, a document usually consists of several topics. Some topics are typically described in detail by more sentences than other topics, and hence may be inferred to comprise the major (or most important) content of the document. Other topics may be briefly

mentioned to supplement or to support the major topics, or to make the whole story more complete. Those of skill in the art will appreciate that a good generic text summary should cover the major topics of the document as thoroughly as possible within a prescribed length (word count or sentence count, for example), while at the same time, minimizing redundancy.

(Specification, p. 6, lines 1-6; paragraph [0023]; emphasis added.)

Minimizing redundant sentences regarding the main topic and including sub-topics of a text in the generic text summary are the issues addressed by the present Application.

For example, when a document is about a subject such as hiking, the text of the document is bound to have many sentences about trails. At the same time, it also may have a few sentences about mushrooms and a few other sentences about streams. Because the sentences including the term “trail” are most frequent in the document, a generic or overall summary that is generated for this document would tend to focus on the sentences that include the term “trail” and best represent the document. As a result, the generated summary will include a large number of sentences regarding the “trails.” However, when the size of the summary is set at a prescribed length that is not to be exceeded, then the summary may leave out the other subjects discussed by the document such as the mushrooms and the streams. In other words, the generated summary would be redundant regarding the subject of trails and deficient regarding the subjects of mushrooms and streams. If the summarization method is conducted in steps such that after the sentence or sentences including the term “trail” are determined as the most representative sentences of the document, then these sentences and the word “trail” are deleted from the document, a modified document text would result that does not include the term “trail.” This modified document is used to find the next sentence or sentences to be placed in the summary.

Because the dominant term is now eliminated, the sentences including the mushrooms and the streams will have a chance to be identified and placed in the summary. Without this method, the generic summary is likely to include redundant sentences about trails and no information regarding the mushrooms or streams.

The specification describes the process as follows:

It will be appreciated by those of skill in the art that, at block 104 of the foregoing operation, the sentence, k, having the highest relevance score (relative to the document) may be considered the sentence that best represents the major content of the document. Therefore, selecting sentences based upon relevance scores in the foregoing manner may ensure that tie [sic] summary represents the major topics of the document to the greatest extent possible. On the other hand, eliminating all the terms contained in k from the document, as shown at block 105, may ensure that retrieval of the subsequent sentence (in the following iteration) with the highest relevance score will create minimum overlap with the subject matter contained in sentence k. In this manner, a very low level of redundancy may be achieved during creation of a summary that covers every major topic in the document.

(Specification, p. 8, line 22, to p. 9, line 2; paragraph [0043]; emphasis added.)

Claim 1

Claim 1 recites:

“A method of creating a generic text summary of a document (page 5, lines 28-31); said method comprising:

obtaining the document (page 5, lines 28-31);

creating a weighted document term-frequency vector for said document (page 8, line 1-3; figure 1, element 102);

for each sentence in said document, creating a weighted sentence term-frequency vector (page 8, line 1-3; figure 1, element 102);

computing a score for each said weighted sentence term-frequency vector in accordance with relevance to said weighted document term-frequency vector (page 8, line 3-6; figure 1, element 103);

selecting a sentence for inclusion in said generic text summary in accordance with said computing, wherein the selected sentence has the computed score representing high degree of relevance of the corresponding weighted sentence term-frequency vector to said weighted document term-frequency vector (page 8, line 3-6; figure 1, element 104);

deleting said selected sentence from said document and eliminating terms in said selected sentence from said document (page 8, line 12-17; figure 1, element 105); and

generating the generic text summary based on the selected sentence. (figure 1, elements 103, 104, 105, page 8, line 22 - page 9, line 2);”

Claim 9

Claim 9 recites:

“A system for creating a generic text summary of a document (page 5, lines 28-31); said system comprising:

a computer comprising at least a CPU and a memory;

an interface for obtaining the document (page 5, lines 28-31);

a display for displaying said generic text summary (page 5, lines 28-31); and

summarizer program code (page 5, lines 28-31), operable on said computer, for analyzing and summarizing said document; said summarizer program code comprising:

a vector generator for creating a weighted document term-frequency vector for said document and creating a weighted sentence term-frequency vector for each sentence in said document (page 8, line 1-3; figure 1, element 102);

a scoring engine for computing a score for each said weighted sentence term-frequency vector in accordance with relevance to said weighted document term-frequency vector (page 8, line 3-6; figure 1, element 103);

a selector for selecting a sentence for inclusion in said generic text summary in accordance with output results from said scoring engine (page 8, line 3-6; figure 1, element 104);

a document editor for deleting said selected sentence from said document and for eliminating terms in said selected sentence from said document (page 8, line 12-17; figure 1, element 105) and;

a generic summary generator for generating the generic text summary based on the selected sentence. (figure 1, elements 103, 104, 105, page 8, line 22 - page 9, line 2)”

Claim 13

Claim 13 recites:

“A method of creating a generic text summary of a document (page 5, lines 28-31); said method comprising:

obtaining the document (page 5, lines 28-31);

decomposing said document into individual sentences (figure 1, element 101);

forming a candidate sentence set from said individual sentences (figure 1, element 101);

for each of said individual sentences in said candidate sentence set, creating a weighted sentence term-frequency vector (page 8, line 1-3; figure 1, element 102);

creating a weighted document term-frequency vector for said document (page 8, line 1-3; figure 1, element 102);

for each of said individual sentences in said candidate sentence set, computing a relevance score for said weighted sentence term-frequency vector relative to said weighted document term-frequency vector (page 8, line 3-6; figure 1, element 103);

selecting a sentence for inclusion in said generic text summary in accordance with said computing, wherein the selected sentence has the computed relevance score representing a high degree of relevance of the corresponding weighted sentence term-frequency vector to said weighted document term-frequency vector (page 8, line 3-6; figure 1, element 104);

deleting said selected sentence from said candidate sentence set (page 8, line 12-17; figure 1, element 105);

eliminating terms in said selected sentence from said document (page 8, line 12-17; figure 1, element 105);

recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating (page 8, line 12-17; figure 1, element 105); and

generating the generic text summary based on the selected sentence (figure 1, elements 103, 104, 105, page 8, line 22 - page 9, line 2)."

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

This Appeal is directed to all rejections of the claims 1-20 that are under Appeal. In particular, the grounds of rejection to be reviewed include:

1- Whether Claims 1-4 and 9-16 are unpatentable under 35 U.S.C. 103(a) over Hwang (U.S. Patent Application Publication No. 2002/0078090 A1) in view of Foltz (U.S. Patent No. 6,356,864)?

2- Whether Claims 5, 7, 17 and 19 are unpatentable under 35 U.S.C. 103(a) over Hwang, in view of Foltz, further in view of Boguraev (U.S. Patent No. 6,865,572)?

3- Whether Claims 6, 8, 18 and 20 are unpatentable under 35 U.S.C. 103(a) over Hwang, in view of Foltz, further in view of Boguraev, and further in view of Cohen (William W. Cohen, "Data Integration Using Similarity Joins and a Word-Based Information Representation Language," by AT&T Lab-Research Shannon Laboratory, Published by ACM, Vol. 18; No. 3, July 200, pages 288-321.")?

VII. ARGUMENT

1. Whether Claims 1-4 and 9-16 are unpatentable under 35 U.S.C. 103(a) over Hwang in view of Foltz?

1(a). Claim 13

Regarding claim 13, Appellant submits that at least the following elements are not taught or suggested by the references:

- “deleting said selected sentence from said candidate sentence set;”
- “eliminating terms in said selected sentence from said document;” or
- “recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating.”

Hwang and Foltz are the references cited for allegedly teaching the elements of claim 13.

Hwang, and more specifically the act of “extraction” of sentences to generate the summary in Hwang, is cited for allegedly teaching the first two of the above elements of the claim. The “extraction” of Hwang in combination with its processing of new documents are cited for allegedly teaching the third element, above. Appellant submits, the “extraction” in Hwang does not teach or suggest “selecting a sentence for inclusion in said generic text summary ... [and] deleting said selected sentence from said candidate sentence set” or “eliminating terms in said selected sentence from said document” and there is no disclosure of “recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating” in Hwang.

Foltz and the other cited references do not cure the deficiency of Hwang.

Hwang

Claim 13 recites:

“A method of creating a generic text summary of a document; said method comprising:

obtaining the document;

decomposing said document into individual sentences;

forming a candidate sentence set from said individual sentences;

for each of said individual sentences in said candidate sentence set, creating a weighted sentence term-frequency vector;

creating a weighted document term-frequency vector for said document;

for each of said individual sentences in said candidate sentence set, computing a relevance score for said weighted sentence term-frequency vector relative to said weighted document term-frequency vector;

selecting a sentence for inclusion in said generic text summary in accordance with said computing, wherein the selected sentence has the computed relevance score representing a high degree of relevance of the corresponding weighted sentence term-frequency vector to said weighted document term-frequency vector;

deleting said selected sentence from said candidate sentence set;

eliminating terms in said selected sentence from said document;

recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating; and

generating the generic text summary based on the selected sentence.”

(Emphasis added.)

As set forth above, Hwang is cited by the Office action for allegedly teaching or suggesting the underlined elements of the claim. Appellant demonstrates below that Hwang does not teach or suggest the above elements.

Hwang generates a query-relevant summary and is not directed to “a method of creating a generic text summary” of the claims.

Hwang is cited for allegedly teaching all of the elements of the claim except the element of “creating a weighted document term-frequency vector” and the related elements. (See, Office action of September 4, 2008, pp. 4-6 and p. 7.)

Appellant submits that Hwang is a query-based text summarization method that generates a query-relevant summary. Query-relevant summaries are distinguished from the generic text summaries in the “background of the invention” section of the present Application:

[0008] There are two types of text summaries: generic summaries, and query-relevant summaries. Generic summaries provide an overall sense of a particular document's content, while query-relevant summaries present only content from a particular document that is closely related to the initial search query.

[0009] A good generic summary should contain the main topics presented in a document while minimizing redundancy. Since the generic summarization process is not responsive to a particular keyword query or topic search, developing a high quality generic summarization method and system has proven very challenging. A query-relevant summary, on the other hand, presents document contents that are specifically related to an initial search query; in many existing systems, creating a query-relevant summary is essentially a process of retrieving query-relevant sentences from the document. It will be appreciated by those of skill in the art that this process is strongly related to the text retrieval process. Accordingly, query-relevant summarization is most often achieved simply by extending conventional IR technologies.

(Gong, U.S. Patent Application Publication No. 2002/0138528; paragraphs [0008] and [0009]; emphasis added.)

Hwang is directed to “text summarization using relevance measures and latent semantic analysis.” (Hwang, Title.) The abstract of Hwang states:

At least one domain ontology that includes a set of concepts is selected. A user profile indicative of a user's interests is defined in terms of the ontology concepts. A document's relevance to the user is determined based upon the user profile. If the document is relevant, at least a portion of the ontology is used to extract concepts from the document. The degree of match between the extracted concepts and the user profile concepts is determined and the document text summary is generated if the degree of match exceeds a predetermined threshold. Generating the summary may include selecting sentences based on the concepts in the user profile, ranking the selected sentences by relevance to the user profile, selecting sentences for inclusion in the document text summary based upon the ranking, and merging the selected sentences into the document text summary.

(Hwang, Abstract; emphasis added.)

Figure 3 of Hwang is “a flow diagram depicting a detailed method of constructing the summarization process. (Hwang, paragraph [0017].) The flowchart of figure 3 of Hwang ties into figure 2 of this reference and corresponds to block 220 of figure 2 that states “generate text summary.” (Hwang, paragraph [0034].) The text summarization flowchart of figure 3 begins by a step 302 of “select sentences with desired concept.” According to the written description of this drawing: “In one embodiment, all sentences in the original document that contain concept terms that would interest the user ... are marked for selection.” (Hwang, paragraph [0036].) As the above citations to Hwang indicate, Hwang generates a text-summary according to relevance of a sentence to an external factor such as “user interest” and NOT according to relevance of this

sentence to the document itself which determines whether the sentence is representative of the document. Therefore, unlike the present Application, Hwang is directed to a query-based text summarization method.

After step 302, in step 304 “additional sentences are marked as candidates to be included in the summary.” (Hwang, paragraph [0037].) These additional sentences are marked based on context. (See figure 3, step 304.) This means that, for example, if the sentence selected in step 302 includes a pronoun and the antecedent of the pronoun is not clear from the candidate sentence of step 302, then in step 304, the sentence prior to the sentence of 302 is also marked in order to provide antecedence for the pronoun in the sentence selected at 302. This way, the summary will make sense and will not include pronouns that are unresolved for the reader. Later, at 308, a relevance score for each sentence is computed. (Hwang, paragraph [0039].) “The relevance score may be based on several factors including conceptual relevancy ... thematic relevancy ... and the probability that a particular sentence may contain the antecedent of unresolved anaphors.” (Hwang, paragraph [0039].) The relevance that is determined in step 308, is the relevance of the sentence with an external subject or question; it is not a relevance between the sentence and the document that is the source of the sentence. Next at 310 and 312, the sentences that have been selected as relevant are ranked and the higher ranking sentences are selected for being included in the summary. (Hwang, figure 3, and paragraph [0040].) At 314, the sentences determined for inclusion are extracted along with their context information, such as the paragraph they are from (paragraph [0041], and at 316, the extracted sentences are merged into the text summary. (Hwang, paragraph [0041].) Due to the query-relevant nature of the summary generated by Hwang, all of the sentences selected for being included in the summary of

Hwang are extracted at the same time at step 314 or can be extracted at the same time without impacting the resulting summary.

The claim element of “deleting said selected sentence from said candidate sentence set” is not taught or suggested by Hwang.

The abstract of Hwang and paragraphs [0040] and [0041] of this reference are cited for teaching the elements of the claim. (See Office action of September 4, 2008, pp. 3-4 and p. 7.) The abstract was provided above. The cited paragraphs that describe the later steps, 310-316, of the flowchart of figure 3 of Hwang are provided below:

[0040] The selected sentences are then ranked (block 310) by their score. Based upon the ranking of the sentences and a pre-defined criteria, the sentences that are to be included in the summary are determined in block 312. In one embodiment, the length of the proposed summary, whether user selected or automatically generated, is taken into account in deciding which sentences are to be included. In this embodiment, the score a sentence must achieve before being selected for inclusion in the text summary increases as the desired length of the summary decreases.

[0041] The sentences determined for inclusion are then extracted (block 314) along with any desired context information (e.g., which paragraph each sentence is from, etc.) and merged. If the number of sentences is large enough, the sentences may be grouped into two or more paragraphs. Paragraph break points are then determined (block 316) based upon the interdependency between the sentences in the merged text to form paragraphs in the text summary. (Hwang, paragraphs [0040] and [0041]; emphasis added.)

The Examiner is interpreting the “extraction” of sentences in Hwang as the “deleting” of the claim. (See, e.g., Advisory Action of December 18, 2008, p. 2, 4th paragraph, line 4.) Hwang extracts the sentences to form a summary of the text. However, Hwang does not delete

the sentences, which have been extracted, from the document. The text summarization of Hwang is a query-based text summarization method that is explained as an alternative to the “generic text summarization” that is the subject of the Application. (See, e.g., paragraph [0009] of the published Application (US 2002/0138528).) All of the sentences that are used to form the summary are determined by referring to the same unaltered and unmodified document (step 312 of figure 3 of Hwang) and looking for sentences that are relevant to a particular subject or query and subsequently merging the relevant sentences into the summary (step 314 of figure 3 of Hwang).

Therefore, **first**, there is no disclosure in Hwang that the “extracted” sentences are deleted from the document; **second**, because the summary of Hwang is generated in one step, there is no reason for deletion of the selected sentences in Hwang; and even if the summary of Hwang were generated in several consecutive rounds, deleting the extracted sentence from the document would be a meaningless extra step. These two points are further discussed below.

First, there is no disclosure in Hwang that the “extracted” sentences are deleted from the document. (See Office action of September 4, 2008, p. 3, 3rd line from the end.) “Extracting” and “deleting” are not the same. If “extraction” were followed by “discarding the extracted item,” and were intended to mean “exclusion,” then maybe “extraction” could be “deletion.” Here, the Examiner is using “extracting” to indicate “inclusion” in the text summary (see page 3 of the Office action of September 4, 2008: “Hwang discloses this limitation in that selecting sentences for inclusion in the document text summary based upon the ranking”). Therefore, “extracting” cannot mean exclusion at the same time that is used to mean “inclusion” and is not

synonymous with “deletion.” The Examiner is interpreting the “extraction” of Hwang to have several contradictory meanings at the same time.

In the claim, the sentences that are “deleted” happen to be the sentences that were “selected” for inclusion in the summary. Yet, “deleting” of the claim is an element in addition to elements of “selecting a sentence for inclusion in ... summary” and “generating the generic text summary based on the selected sentence.” The claim gets rid of the sentences that have already been used in the summary, and in doing so modifies the document, so it can calculate the next relevance score unaffected by the sentences that have already been used. The reference Hwang, on the other hand, appears to be selecting the sentences that are relevant to user’s interest, for example, and extracting them for inclusion in the summary; extraction of Hwang could be copying the sentence and pasting it in a summary. It is not deletion of the sentence from the original document.

Second, there is no reason for deletion of the selected sentences in Hwang. Deletion of a selected sentence changes the document before a subsequent selection step. Hwang teaches that the sentences are selected all together at step 312 of figure 3; at step 314 of figure 3, the selected sentences are merged into a summary. Later, at steps 316 and 318, this summary is further reorganized and refined but no more sentences are selected or added to it. Under this scenario, after the sentences are selected and included in the summary, the interaction with the document ends and there is no point in modifying the document by deleting the selected sentences from it.

Further, Hwang is determining the relevance of each sentence to a “user profile indicative of user’s interests” which is a factor unrelated to the document. (Hwang, abstract.) When the relevance of each sentence is being gauged with respect to factors external to the document as is

done in Hwang, there is no reason for modifying the document by deleting a sentence from the document.

The element of “eliminating terms in said selected sentence from said document” of the claim is not taught or suggested by Hwang.

Eliminating terms in the selected sentence from the document is either a futile extra step or renders Hwang inoperable for its intended purpose. Hwang appears to develop its summary by selecting and extracting all the pertinent sentences at once. As such, there is no point in modifying the document by eliminating the terms of the selected sentences from the document after the task of generating the summary is complete. However, if Hwang operated iteratively by referring to the document, selecting some sentences for the summary, extracting these sentences, referring back to the document a second time and selecting and extracting some more sentences, then eliminating the terms in the extracted sentences from the document, after each round, would render Hwang inoperable. This point may be illustrated by returning to the example of the hiking document that had frequent mentions to “trail” as well as infrequent occurrences of “mushrooms” and “streams.” If the user of Hwang is interested in the subject of trails, he would appreciate obtaining a summary that includes all the important material about the trails. This user does not care what the document says “in general;” he does not care what the document says about “mushrooms;” he only cares what the document says about the “trails.” If Hwang proceeds to delete the sentence about the trails and further remove all mentions of the term “trail” from the document, the next round of summarization would return no sentences regarding the requested subject of “trails.” The summarization method would work for one round and would stop. Further, depending on the ranking and selection criteria, the generated summary

may not reflect all the material pertinent to trails. Eliminating the terms that were on point before running another round of search on the document, subverts the objective of Hwang. Therefore, if Hwang were to generate its summary in successive rounds of selection and extraction, then deleting sentences or eliminating the terms of those sentences would effectively stop the operation after the first round.

The Examiner appears to be citing to “context information e.g., which paragraph each sentence is from” of Hwang for teaching “eliminating terms in said selected sentence” of the claim. (See Office action of September 4, 2008, p.3.) There are two places that “context” appears in Hwang: (1) in step 304 of figure 3 where “context-charged” sentences that, for example, provide pronoun resolution are marked in addition to the sentences selected in step 302; and (2) in step 314 where “the sentences determined for inclusion are then extracted (block 314) along with any desired context information (e.g., which paragraph each sentence is from, etc.) and merged.” (Hwang, paragraphs [0037] and [0041].) Neither meaning of “context” in Hwang teaches the element of “eliminating terms” of the claim.

If the terms in the “context-charged” sentences of step 302 of figure 3 of Hwang are eliminated, then as mentioned above, the “eliminating” of such terms is (a) meaningless if all the summary sentences are extracted at once or (b) it would render Hwang inoperable for its intended purpose if the summary sentences are extracted in consecutive turns.

According to the hiking document example, if the user is interested in obtaining a summary of the document that is focused on the subject of trails, deleting all mentions to the term “trail,” after the higher ranking sentences including the phrase “trail” have been selected, is either (a) meaningless or (b) defeats his purpose. If the summary is generated after one round of

selection and extraction; then the process is done and there is no point in modifying the document by deleting some of its sentences or eliminating the terms of these sentences from the document, hence situation (a). If the summary is generated in several rounds of selection and extraction, then the summary that would be generated would include only those sentences regarding “trails” that were extracted after the first round and then because the term “trail” is eliminated from the document, the subsequent rounds of selection cannot find any further sentences including this term, hence situation (b).

The element of “recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating” is not taught or suggested by Hwang.

The Examiner adds paragraph [0047] of Hwang that pertains to “Processing new documents against pre-selected, client specific concepts defined by the client, or inferred by the system, and computing the relevancy score for each document” for teaching the above element. (See, Office action of September 4, 2009, pp. 11, 12, in the context of rejection of claim 2; emphasis added.) Paragraph 47 of Hwang is provided below:

[0047] For each client, when a new document arrives, the system checks if the document is relevant to the client. Processing new documents against pre-selected, client-specific concepts defined by the client, or inferred by the system, and computing the relevancy score for each document, the system can perform a continual text summarization method. The relevance score is computed based on several factors, such as the number of ontological concepts found in the document that match (or are associated with) the pre-selected, client-specific concepts (in case of associated concepts), the strength of the concept (i.e., the inverse of the distance on the ontology between the interesting-concept and the corresponding concept found in the document), the number of matches, etc. If the relevance of the document is larger than a user defined threshold, the system extracts the relevant concepts together with the sentences, or a region of sentences such

as paragraphs, in which they occur. The system then determines the themes running through the extracted portion of the document. Words and phrases whose frequencies yield high with respect to their prior probabilities are selected as themes. Themes do not have to be ontological concepts.

(Hwang, paragraph [0047]; emphasis added.)

The above paragraph of Hwang refers to “continual text summarization” by continually moving to “new documents” and is not addressing continual summarization of the text of the same first document. Therefore, the element of “recreating said weighted document term-frequency vector” of the claim appears to be interpreted by the Examiner as pertaining to a new document. This is incorrect. The element sets forth “recreating said weighted document term-frequency vector.” This element refers back to “creating a weighted document term-frequency vector for said document” of the claim and pertains to the same document that was processed earlier in the claim. Hwang has no disclosure of repeating the process on the same document after “said deleting and said eliminating.” Rather, in Hwang it is one document, one time and then onward to the next document. As such, Hwang cannot teach “deleting ... eliminating ... and ... recreating ... in according with ... deleting and ... eliminating” of the claim. Further, as explained above, repeating the process of “deleting and... eliminating,” on the same document is either pointless or renders Hwang inoperable for generating a query-relevant summary.

Foltz

Foltz is cited for allegedly teaching the “creating a weighted sentence term-frequency vector” and does not cure the deficiency of Hwang. (Office action of September 4, 2008, pp. 4-6 and 7-8.) This reference is directed to text evaluation and not to text summarization. Accordingly, Foltz does not cure the deficiency of Hwang with respect to teaching the above-discussed elements of the claim.

Accordingly, Claim 13 is believed to be patentable in view of Hwang and Foltz, whether considered alone or in combination.

1(b). Claim 1

Independent claim 1 includes the elements of “deleting said selected sentence from said document and eliminating terms in said selected sentence from said document,” which is not taught or suggested Hwang and Foltz, whether considered alone or in combination, as described above with reference to claim 13. Accordingly, claim 1 is believed to be patentable in view of Hwang and Foltz, whether considered alone or in combination.

1(c). Claim 9

Independent claim 9 includes the element of “a document editor for deleting said selected sentence from said document and for eliminating terms in said selected sentence from said document,” which is not taught or suggested Hwang and Foltz, whether considered alone or in combination, as described above with reference to claim 13. Accordingly, claim 9 is believed to be patentable in view of Hwang and Foltz, whether considered alone or in combination.

1(d). Claims 2-4, 10-12 and 14-16

Claims 2-4 depend from claim 1. Claims 10-12 depend from claim 9. Claims 14-16 depend from claim 13.

These dependent claims are believed to be patentable at least by virtue of dependence from their base independent claims.

2. Whether Claims 5, 7, 17 and 19 are unpatentable under 35 U.S.C. 103(a) over Hwang, in view of Foltz, further in view of Boguraev?

Boguraev was cited by the Office action for allegedly teaching the “weighting functions” used in some of the dependent claims. (Office action of September 4, 2008, p. 19.) This additional reference does not cure the deficiency of Hwang and Foltz in teaching or suggesting the representative claim 13. The representative independent claim 13, therefore, remains patentable over all cited references Hwang, Foltz and Boguraev whether alone or in combination.

Independent claims 1 and 9 are patentable for reasons similar to those presented regarding claim 13.

Claims 5 and 7 depend from claim 1. Claims 17 and 19 depend from claim 13.

These dependent claims are believed to be patentable at least by virtue of dependence from their respective base independent claims.

3. Whether Claims 6, 8, 18 and 20 are unpatentable under 35 U.S.C. 103(a) over Hwang, in view of Foltz, further in view of Boguraev, and further in view of Cohen?

Cohen is cited for allegedly teaching the normalizing of the vectors. (Office action of September 4, 2008, p. 22.) However, the cited portions of this reference do not appear to cure the deficiencies of Hwang, Foltz and Boguraev in teaching or suggesting the elements of the independent claim 13. The representative independent claim 13, therefore, remains patentable over all cited references Hwang, Foltz, Boguraev and Cohen, whether alone or in combination.

Independent claims 1 and 9 are patentable for reasons similar to those presented regarding claim 13.

Claims 6 and 8 depend from claim 1. Claims 18 and 20 depend from claim 13.

These dependent claims are believed to be patentable at least by virtue of dependence from their base independent claims.

Conclusion

The USPTO is directed and authorized to charge the statutory fee (37 C.F.R. §41.37(a) and 1.17(c)) and all required fees, except for the Issue Fee and the Publication Fee, to Deposit Account No. 19-4880. Please also credit any overpayments to said Deposit Account.

Respectfully submitted,

A handwritten signature in dark ink, appearing to read 'Fariba Sirjani', is written over a horizontal line.

Fariba Sirjani
Registration No. 47,947

SUGHRUE MION, PLLC
Telephone: (202) 293-7060
Facsimile: (202) 293-7860

WASHINGTON DC OFFICE

23373

CUSTOMER NUMBER

Date: April 3, 2009

CLAIMS APPENDIX

CLAIMS 1-20 On Appeal:

1. (rejected) A method of creating a generic text summary of a document; said method comprising:

obtaining the document;

creating a weighted document term-frequency vector for said document;

for each sentence in said document, creating a weighted sentence term-frequency vector;

computing a score for each said weighted sentence term-frequency vector in accordance with relevance to said weighted document term-frequency vector;

selecting a sentence for inclusion in said generic text summary in accordance with said computing, wherein the selected sentence has the computed score representing high degree of relevance of the corresponding weighted sentence term-frequency vector to said weighted document term-frequency vector;

deleting said selected sentence from said document and eliminating terms in said selected sentence from said document; and

generating the generic text summary based on the selected sentence.

2. (rejected) The method of claim 1 further comprising:

recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating; and

selectively repeating said computing, said selecting, said deleting, said eliminating, and said recreating.

3. (rejected) The method of claim 2 wherein said selectively repeating is terminated when a predetermined number of sentences has been selected.

4. (rejected) The method of claim 1 wherein said computing comprises calculating an inner product of said weighted sentence term-frequency vector and said weighted document term-frequency vector.

5. (rejected) The method of claim 1 wherein said creating a weighted sentence term-frequency vector comprises implementing a local weighting function and implementing a global weighting function.

6. (rejected) The method of claim 5 wherein said creating a weighted sentence term-frequency vector comprises normalizing each said weighted sentence term-frequency vector by dividing the weighted sentence term-frequency vector by a magnitude of the weighted sentence term-frequency vector.

7. (rejected) The method of claim 1 wherein said creating a weighted document term-frequency vector comprises implementing a local weighting function and implementing a global weighting function.

8. (rejected) The method of claim 7 wherein said creating a weighted document term-frequency vector comprises normalizing said weighted document term-frequency vector by dividing the weighted document term-frequency vector by a magnitude of the weighted document term-frequency vector.

9. (rejected) A system for creating a generic text summary of a document; said system comprising:

- a computer comprising at least a CPU and a memory;

- an interface for obtaining the document;

- a display for displaying said generic text summary; and

- summarizer program code, operable on said computer, for analyzing and summarizing said document; said summarizer program code comprising:

 - a vector generator for creating a weighted document term-frequency vector for said document and creating a weighted sentence term-frequency vector for each sentence in said document;

 - a scoring engine for computing a score for each said weighted sentence term-frequency vector in accordance with relevance to said weighted document term-frequency vector;

 - a selector for selecting a sentence for inclusion in said generic text summary in accordance with output results from said scoring engine;

 - a document editor for deleting said selected sentence from said document and for eliminating terms in said selected sentence from said document and;

a generic summary generator for generating the generic text summary based on the selected sentence.

10. (rejected) The system of claim 9 wherein said vector generator recreates said weighted document term-frequency vector in accordance with output results from said document editor.

11. (rejected) The system of claim 10 wherein said summarizer further comprises a loop routine for generating iterative sequential operations of said vector generator, said scoring engine, said selector, and said document editor.

12. (rejected) The system of claim 11 wherein said loop routine is responsive to a predetermined limit such that said generic text summary is of a predetermined number of sentences.

13. (rejected) A method of creating a generic text summary of a document; said method comprising:

obtaining the document;

decomposing said document into individual sentences;

forming a candidate sentence set from said individual sentences;

for each of said individual sentences in said candidate sentence set, creating a weighted sentence term-frequency vector;

creating a weighted document term-frequency vector for said document;

for each of said individual sentences in said candidate sentence set, computing a relevance score for said weighted sentence term-frequency vector relative to said weighted document term-frequency vector;

selecting a sentence for inclusion in said generic text summary in accordance with said computing, wherein the selected sentence has the computed relevance score representing a high degree of relevance of the corresponding weighted sentence term-frequency vector to said weighted document term-frequency vector;

deleting said selected sentence from said candidate sentence set;

eliminating terms in said selected sentence from said document;

recreating said weighted document term-frequency vector in accordance with said deleting and said eliminating; and

generating the generic text summary based on the selected sentence.

14. (rejected) The method of claim 13 further comprising:

selectively repeating said computing, said selecting, said deleting, said eliminating, and said recreating.

15. (rejected) The method of claim 14 wherein said selectively repeating is terminated when a predetermined number of sentences has been selected.

16. (rejected) The method of claim 13 wherein said computing comprises calculating an inner product of said weighted sentence term-frequency vector and said weighted document term-frequency vector.

17. (rejected) The method of claim 13 wherein said creating a weighted sentence term-frequency vector comprises implementing a local weighting function and implementing a global weighting function.

18. (rejected) The method of claim 17 wherein said creating a weighted sentence term-frequency vector comprises normalizing each said weighted sentence term-frequency vector by dividing the weighted sentence term-frequency vector by a magnitude of the weighted sentence term-frequency vector.

19. (rejected) The method of claim 13 wherein said creating a weighted document term-frequency vector comprises implementing a local weighting function and implementing a global weighting function.

20. (rejected) The method of claim 19 wherein said creating a weighted document term-frequency vector comprises normalizing said weighted document term-frequency vector by dividing the weighted document term-frequency vector by a magnitude of the weighted document term-frequency vector.

EVIDENCE APPENDIX:

Appellant submits, pursuant to 37 C.F.R. § 41.37(c)(1)(ix), that no evidence submitted pursuant to 37 C.F.R. §§ 1.130, 1.131, or 1.132 or any other evidence, other than the prior art references of record, have been relied upon by Appellant in the Appeal.

RELATED PROCEEDINGS APPENDIX

No proceedings were identified in Section II pursuant to 37 C.F.R. § 41.37(c)(1)(ii). Accordingly, no copies of decisions rendered by a court or the Board in any proceeding identified in Section II pursuant to 37 C.F.R. § 41.37(c)(1)(ii) are submitted herewith.